JOBNAME: lgt 9#1 2005 PAGE: 1 OUTPUT: Sat December 11 02:44:14 2004
lww/lgt/91384/LGT157667
Prod#: LGT157667

Interobserver Agreement for Colposcopy Quality Control Using Digitized Colposcopic Images During the ALTS Trial

Daron G. Ferris, MD,[1] and Mark Litaker, PhD[2] for the ALTS Group
[1]*Gynecologic Cancer Prevention Center, Departments of Family Medicine and Obstetrics and Gynecology, and* [2]*Office of Biostatics and Bioinformatics, The Medical College of Georgia, Augusta, GA*

■ **Abstract**

Objective. To estimate interobserver agreement among colposcopy quality control reviewers viewing digitized cervical images during the Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS).

Materials and Methods. Three colposcopy quality control reviewers independently examined modem-transferred digitized colposcopic images from subjects examined at four clinical centers. Reviewers indicated colposcopic impression, Reid colposcopic index scores, lesion size, and the technical quality of the image. Rates of agreement were evaluated using the κ statistic and McNemar and Bowker tests of symmetry.

Results. Regarding colposcopic impressions, the average weighted κ for pairs of initial reviewers was 0.36 (95% confidence interval, 0.33–0.39). κ scores with respect to Reid colposcopic index, cervical image quality, and lesion size ranged from 0.23 to 0.28, 0.18 to 0.27, and 0.33 to 0.42, respectively.

Conclusions. Fair rates of agreement and poor to fair κ scores among ALTS colposcopy quality control reviewers were noted for colposcopic impression, Reid colposcopic index scores, image quality, and lesion size. Great latitude exists in the interpretation of digitized cervical images. Poor image quality partially may explain these suboptimal results. ■

Key Words: colposcopy, quality control, ALTS, interobserver agreement, cervical neoplasia

Reprint requests to: Daron G. Ferris, MD, Departments of Family Medicine and Obstetrics and Gynecology, 1423 Harper Street, HH-100, Medical College of Georgia, Augusta, GA 30912. E-mail: dferris@mail.mcg.edu

Journal of Lower Genital Tract Disease, Volume 9, Number 1, 2005, 29–35

Colposcopists vary considerably with respect to education and training, knowledge, management preferences, diagnostic skill, and clinical experience [1–6]. Colposcopic proficiency directly influences diagnostic accuracy, which ultimately impacts patient management and outcomes. Although this is important for clinical practice, these issues also affect clinical trials that depend heavily on achieving diagnostic precision with respect to cervical neoplasia. Study endpoints, triage thresholds, and inclusion and exclusion criteria demand proper assignment by colposcopists.

Only in the past several years has the inclusion of colposcopy quality control in clinical trials been considered seriously. The Atypical Squamous Cells of Undetermined Significance Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS) was the first National Cancer Institute trial to include colposcopy as part of the overall quality control effort [7]. This colposcopy quality control program assembled a panel of expert colposcopists to review digitized colposcopic images obtained during all trial-related colposcopic examinations [8, 9]. In theory, colposcopy quality control review of each subject's examination was expected to maintain standard care and to improve outcomes. However, little is known about the actual performance of colposcopy quality control, particularly with respect to interobserver variability pertaining to digitized cervical image assessment. The purpose of the current study was to determine the interobserver agreement of the colposcopy quality control group's assessment of digitized cervical images as part of the ALTS trial.

JOBNAME: lgt 9#1 2005 PAGE: 2 OUTPUT: Sat December 11 02:44:15 2004
lww/lgt/91384/LGT157667

MATERIALS AND METHODS

Participants

Three colposcopists under contract by the National Cancer Institute to provide colposcopy quality control for the ALTS trial were considered for this report. Previously, these colposcopists had successfully completed the Colposcopy Recognition Award examination offered by the American Society for Colposcopy and Cervical Pathology [10]. To be included as part of the quality control group, each also performed satisfactorily with respect to pretrial quality control assessment [8]. Their involvement in the ALTS trial was approved by the respective institutional review boards (Medical College of Georgia, University of California, Santa Barbara, and Beth Israel Deaconess Hospital, Boston).

Materials

Quality control participants reviewed all digitized colposcopic images obtained by clinical center colposcopists during the enrollment and follow-up phases of ALTS using standardized personal computers and customized software (DenVu, Tucson, AZ). The software provided a single screen designed to capture all relevant data (Figure 1). These data included a low- and high-magnification cervical image and various colposcopic assessment items. The assessment items consisted of colposcopic adequacy (satisfactory, unsatisfactory), colposcopic impression (normal, cervicitis, atypical metaplasia, low grade, high grade, cancer, vaginal intraepithelial neoplasia, vulvar intraepithelial neoplasia, other), Reid's colposcopic index categories and scores [11, 12], lesion size (one to four quadrants), biopsy intent (yes, no), biopsy location when necessary (a cursor mark measured by pixel and line coordinates), and quality of digitized images (acceptable, out of focus, blood, mucus, etc.). Except for assessment of image quality, clinical center colposcopists entered the same data after examining subjects. Quality control participants received the two modem-transferred cervical images without colposcopists' annotations for assessment. The software automatically coordinated the transfer of images and data by modem from the clinical centers to the appropriate quality control reviewers. The software also was designed to determine agreement of colposcopic impression and biopsy location among the quality control reviewers. As soon as agreement for these two measures was achieved by two of the quality control reviewers, the data were returned by modem to a central computer for archival and safety net purposes.

Study Design

Colposcopic images from each ALTS subject were assessed initially by two colposcopy quality control reviewers, chosen using a simple algorithm. Assessments were performed individually and in a blinded fashion. Except for subject age, the colposcopy quality control reviewers were blinded from all subject data, including referral and pathologic quality control cytologic results. If the initial reviewers agreed as to both colposcopic impression and biopsy site assessments, the data were sent to the central computer for archival purposes. However,

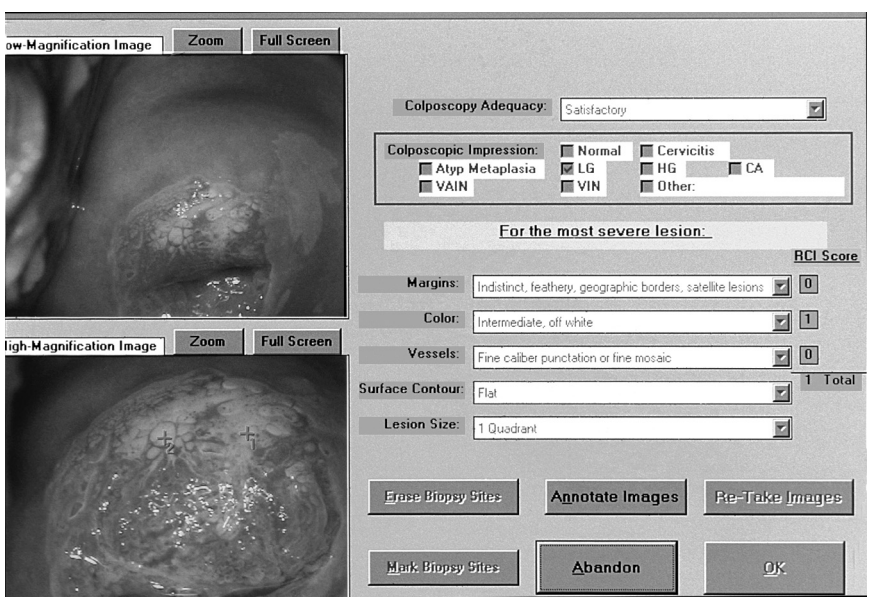

Figure 1. The digital image screen examined by ALTS colposcopy quality control reviewers.

JOBNAME: lgt 9#1 2005 PAGE: 3 OUTPUT: Sat December 11 02:44:16 2004
lww/lgt/91384/LGT157667

if discordance occurred with either interpretation, the images were sent to the third reviewer for adjudication. Concordance was not required for colposcopic adequacy, lesion size, Reid colposcopic index score, or quality of cervical image interpretations. A cervical cancer diagnosis by any reviewer prompted a safety net warning to the clinical center and study coordinator.

Cervical biopsy, endocervical curettage, and electrosurgical loop excision procedures were completed, as necessary, by the clinical center colposcopists. The quality control reviewers did not advise clinicians as to the necessity for these procedures. The histologic specimens were interpreted first by clinical center pathologists and then the ALTS pathology quality control group. For comparative purposes, only the final diagnoses of the pathology quality control group were used. Histologic diagnoses were reported using the CIN classification system.

Statistical Analyses

For each study parameter, degree of agreement was compared between reviewers using κ statistics. To assess systematic differences between raters' interpretations, McNemar and Bowker tests of symmetry were used, depending on whether the variable was dichotomous or polychotomous. Simple percentage agreement statistics were also calculated, with 95% confidence intervals. However, for most of the ratings, a single category was overwhelmingly likely, resulting in a high likelihood of chance agreements between two raters. For this reason, the κ statistics were favored for most assessments. κ scores of 0.40 or less were considered poor, scores between 0.41 and 0.59 were considered fair, and score of 0.60 or more were considered good. Expected agreement rates were calculated using $3 \times 3 \times 3$ tables (SAS PROC FREQ Frequency Procedure; SAS Institute, Cary, NC). Expected counts for the cells representing complete disagreement were added and divided by sample sizes to calculate proportions (percentages). Complements of these percentages were the expected percentages for 2 or 3 of 3 agreement.

RESULTS

There were 4,731 initial colposcopic examinations during ALTS. No quality control colposcopic impressions were available for 1,093 cases because of hardware or software problems associated with our use of this innovative quality control technology, leaving 3,638 evaluable examinations. Colposcopic impression agreement among the colposcopy quality control reviewers based on histologic diagnoses rendered by the pathology quality control group was estimated (Table 1).

All three reviewers' colposcopic impressions disagreed for 12 of 1091 (1.1%) examinations for which there was no biopsy diagnosis, 51 of 1503 (3.4%) examinations with a biopsy diagnosis of normal, 18 of 521 (3.5%) examinations with a biopsy diagnosis of CIN 1, and 29 of 524 (5.5%) examinations with a biopsy diagnosis of CIN 2 or worse. Based on the observed frequencies of each of the diagnostic categories, such three-way disagreement would be expected by chance alone for 1.2% of examinations with no diagnosis, and 5.4%, 5.6%, and 9.0% of examinations with diagnoses of normal, CIN 1, and CIN 2 or worse, respectively. At least two reviewers agreed on 1079 of 1091 (98.9%) examinations for which there was no biopsy diagnosis, 1452 of 1503 (96.6%) examinations with normal biopsy results, 503 of 521 (96.5%) examinations with CIN 1, and 495 of 524 (94.5%) examinations with CIN 2 or worse. If agreement was the result of chance alone, the expected agreement would be 98.8%, 94.6%,

Table 1. Colposcopy Quality Control Pairwise Diagnostic Agreement of Colposcopic Impression by Histologic Results

Quality control reviewers[a]	Colposcopic impression agreement[b] n	%	95% CI	κ	*p* value[c]
Histologic diagnosis: none					
1 and 2	898/1091	82.3	(79.9, 84.5)	0.21	0.4099
1 and 3	363/510	71.2	(67.0, 75.1)	0.09	0.0733
2 and 3	358/510	70.2	(66.0, 74.1)	0.14	0.0003[d]
Histologic diagnosis: normal					
1 and 2	940/1503	62.5	(60.0, 65.0)	0.25	0.8986
1 and 3	631/1110	56.8	(53.9, 59.8)	0.23	0.0002[e]
2 and 3	700/1110	63.1	(60.1, 65.9)	0.35	0.0002[f]
Histologic diagnosis: CIN 1[g]					
1 and 2	291/521	55.9	(51.5, 60.2)	0.29	0.0635
1 and 3	205/422	48.6	(43.7, 53.5)	0.20	0.0733
2 and 3	255/422	60.4	(55.6, 65.1)	0.38	0.0066[h]
Histologic diagnosis: ≥CIN 2					
1 and 2	265/524	50.6	(46.2, 54.9)	0.32	0.0266[i]
1 and 3	252/456	55.3	(50.6, 59.9)	0.39	0.1319
2 and 3	250/456	54.8	(50.1, 59.5)	0.38	0.0012[j]

[a]Pairs of colposcopy quality control reviewers.
[b]Number, percent agreement, and 95% CI (confidence interval) for colposcopic impression agreement with histologic results.
[c]Bowker test of symmetry.
[d]Reviewer 2 tended to rate higher than reviewer 3.
[e]Reviewer 1 rated more disagreements as CIN 1 vs. normal than reviewer 3.
Reviewer 3 rated more disagreements as ≥CIN 2 vs. normal or CIN 1 reviewer 1.
[f]Reviewer 3 rated more disagreements as ≥CIN 2.
[g]Cervical intraepithelial neoplasia.
[h]Reviewer 2 rated more disagreements as CIN 1 vs. normal than reviewer 3.
Reviewer 3 rated more disagreements as ≥CIN 1 vs. normal or CIN 2 than reviewer 2.
[i]Reviewer 2 tended to rate higher than reviewer 1.
[j]Reviewer 3 rated more disagreements as CIN 1 than did reviewer 2.
Reviewer 2 rated more disagreements as CIN 1 than did reviewer 3.
Reviewer 3 diagnosed ≥CIN 1 when reviewer 2 diagnosed normal in 49 cases, whereas reviewer 2 diagnosed ≥CIN 1 when reviewer 3 diagnosed normal in 89 cases.

94.4%, and 91.0% for no diagnosis, normal, CIN 1, and CIN 2 or worse, respectively. For all diagnoses, weighted κ for the initial two reviewers was 0.36 (95% confidence interval, 0.33–0.39). Other agreement and κ values for each pair of reviewers are presented in Table 1. Highest colposcopic impression κ scores (0.32–0.39) were noted for subjects with histologic diagnoses of CIN 2 or worse.

We also estimated pairwise agreement of reviewers with respect to Reid colposcopic index scores (Table 2). The rates of agreement varied between 84.0% and 89.3%, reflecting the high percentage of scores of 0 to 2. κ scores accordingly were much lower, ranging from 0.23 to 0.28. Significant scoring differences were observed between reviewers 1 and 3 ($p = 0.001$), and reviewers 2 and 3 ($p = 0.002$) because the third reviewer tended to derive higher scores than the first two reviewers.

Table 3 depicts colposcopy quality control reviewers' pairwise agreement of digitized cervical image technical quality. Quality agreement ranged between 70.0% and 73.0%. κ scores varied from 0.18 to 0.27. Reviewer 1 tended to rate a significantly greater number of images as acceptable compared with the other two reviewers. All three reviewers rated 1,403 of 2,471 (56.8%; 95% confidence interval, 54.8–58.7) images as adequate.

Agreement of lesion size by reviewers also was estimated (Table 4). Of 3,639 examinations, two-way agreement was reported for 2,593 (71.3%) cases, three-way agreement for 463 (12.7%), and no agreement for 582 (16.0%). Pairwise agreement of lesion size ranged between 42.3% and 50.6%. κ values varied from 0.33 to 0.42. Significant differences were noted for all pairwise comparisons as greater lesion size was indicated by reviewer 2 > 1 > 3.

Table 2. Colposcopy Quality Control Pairwise Diagnostic Agreement of Reid Colposcopic Index Scores[a]

Quality control reviewers[b]	Reid colposcopic index agreement[c]			κ[d]	p value[e]
	n	%	95% CI		
1 and 2	3248/3638	89.3	(88.2, 90.3)	0.23	0.09
1 and 3	2115/2497	84.7	(83.2, 86.1)	0.28	0.001[f]
2 and 3	2098/2498	84.0	(82.5, 85.4)	0.25	0.002[g]

[a]Reid colposcopic index categories of margin, color, and vessels were considered. Total scores grouped as 0–2, 3, and 4–6. Iodine scores were not included. A score of 0–2 is predictive of immature metaplasia or CIN 1,3 equates to CIN 1 or 2, and 4–6 is CIN 2,3.
[b]Pairs of colposcopy quality control reviewers.
[c]Number, percent agreement, and 95% CI (confidence interval) for Reid colposcopic index.
[d]Weighted κ.
[e]Bowker test of symmetry.
[f]More disagreement in the direction of reviewer 3 > reviewer 1 than in the opposite direction.
[g]More disagreement in the direction of reviewer 3 > reviewer 2 than in the opposite direction.

Table 3. Colposcopy Quality Control Reviewers' Pairwise Agreement of Digitized Cervical Image Technical Quality

Quality control reviewers[a]	Adequacy of cervical image agreement[b]			κ	p value[c]
	n	%	95% CI		
1 and 2	2651/3631	73.0	(71.5, 74.4)	0.27	<0.0001[d]
1 and 3	1752/2471	70.9	(69.1, 72.7)	0.18	<0.0001[e]
2 and 3	1729/2471	70.0	(68.1, 71.8)	0.23	0.04[f]

[a]Pairs of colposcopy quality control reviewers.
[b]Number, percent agreement, and 95% CI (confidence interval) of digitized cervical image technical quality. Unacceptable classifications include obsuring blood, prolapsing vaginal sidewalls, poorly focused, unable to see entire squamocolumnar junction, unable to see complete cervical lesion if present, and obsuring mucus.
[c]McNemar test.
[d]Reviewer 1 rated more images as adequate than reviewer 2.
[e]Reviewer 1 rated more images as adequate than reviewer 3.
[f]Reviewer 3 rated more images as adequate than reviewer 2.

CONCLUSIONS

Visual interpretation of infinitely varied objects thought to be classified into several distinct categories requires great skill. Classification of the spectrum of cervical epithelial features as unique histologic equivalents of subjective colposcopic findings has become routine practice during the management of women with lower genital tract neoplasia. We examined the ability of a colposcopy quality control group to derive identical assessments of digitized cervical images obtained during the ALTS trial. Although interobserver rates of agreement for colposcopic impression were fair, the κ values were poor. In a previous publication, we reported rather low rates of colposcopic impression agreement with histologic diagnoses rendered by the ALTS pathology quality

Table 4. Colposcopy Quality Control Reviewers' Agreement of Cervical lesion Size Based on Cervical Quadrant Involvement

Quality control reviewers[a]	Lesion size (quadrants) agreement[b]			κ[c]	p value[d]
	n	%	95% CI		
1 and 2	1839/3638	50.5	(48.9, 52.1)	0.42	<0.0001[e]
1 and 3	1055/2498	42.2	(40.3, 44.2)	0.33	<0.0001[f]
2 and 3	1090/2498	43.6	(41.7, 45.6)	0.38	<0.0001[g]

[a]Pairs of colposcopy quality control reviewers.
[b]Number, percent agreement, and 95% CI (confidence interval) for lesion size rated as quadrants (1–4).
[c]Weighted κ.
[d]Bowker test of symmetry.
[e]More disagreements were reviewer 2 > reviewer 1 than in the opposite direction.
[f]More disagreements were reviewer 1 > reviewer 3 than in the opposite direction.
[g]More disagreements were reviewer 2 > reviewer 3 than in the opposite direction.

JOBNAME: lgt 9#1 2005 PAGE: 5 OUTPUT: Sat December 11 02:44:18 2004
lww/lgt/91384/LGT157667

control group [9]. The reason for poor colposcopic impression agreement among the colposcopy quality control group may be the result of the inherent subjective process involved and the difficulty of evaluating digitized cervical images. This is particularly true when the quality of the image is less than optimal, as was the case for many digitized ALTS cervical images. Furthermore should be remembered that our results represent assessment of static digitized cervical images and not those results expected when actually examining a patient with a colposcope.

The colposcopy quality control group demonstrated fair rates of agreement (84.0% to 89.3%) with respect to Reid colposcopic index scores. κ values (0.23–0.28) were poor, primarily because of the large number of subjects with minimal cytologic changes and a normal cervix, relative to the number with disease, and to one reviewer who tended to derive higher scores. Agreement may have improved if the fourth component of the Reid colposcopic index, iodine staining, was included in the trial. A faint acetowhite lesion with a fine vascular pattern may have been classified more accurately by an additional assessment after Lugol's iodine application [11, 12]. Critical examination of the groups' detailed assessments based on typical colposcopic signs (lesion color, margin, contour, vessels, etc.) may help explain tendencies to over- or underdiagnose.

Modest interobserver agreement is not unique to colposcopy. Histologic and cytologic diagnoses vary among pathologists because of the similar subjective task. The interobserver agreement for cytologic and histologic assessment in the ALTS trial measured by the κ statistic was fair (κ = 0.46) [13]. Other studies of interobserver agreement of histologic diagnoses by pathologists have reported κ values ranging from 0.15 to 0.55 [14, 15]. Comparable levels of agreement are seen for colposcopists with regard to assessment of colpophotographs and cervigrams [8, 16, 17]. Interobserver agreement of colposcopic diagnoses formulated after review of colposcopy videotapes was poor (κ = 0.17) in a study from the United Kingdom [18]. We previously reported the agreement of colposcopic experts' diagnoses viewing real time videorecorded colposcopic examinations to be fair (64%; κ = 0.42) [19]. This level of interobserver agreement was similar to that of reviewing digitized colposcopic images (65.2%; κ = 0.43) as part of a comparison of two distinct types of telecolposcopy [20]. The reason for the better agreement in the aforementioned study compared with this current analysis may result from a different group of reviewers, knowledge of the referral Pap smear result, and a subject population with a greater spectrum of cytologic diagnoses. Therefore, diagnostic agreement among pathologists and clinicians varies substantially whether examining at the microscopic or more macroscopic level.

As part of the quality control effort, reviewers assessed the technical quality of digitized cervical images. Colposcopists received remedial training if found to exceed the established threshold for technically unsatisfactory images. Reviewers had poor κ rates when assessing image quality. This assessment included diverse reasons for inadequate images: poorly focused or too dark, obscuring blood mucus or vaginal sidewalls, and inability to visualize the entire squamocolumnar junction. The reported rates of agreement are considered somewhat disappointing, but this assessment also was very subjective and reviewers may have inconsistently tolerated mildly to moderately poor technical quality.

The size of cervical lesions correlates directly with the severity of disease. The ALTS colposcopy reviewers demonstrated poor to fair κ scores and fair agreement (71.3%) for estimation of lesion size. Hopman et al. [17] reported an interobserver agreement rate of 68% when evaluating colposcopic photographs for lesion size. Size was determined by quadrants of the cervix involved, as is the case clinically. The ALTS reviewers also classified lesion size based on the number of quadrants involved. However, such an assessment may be arbitrary because a very small, linear lesion may be present in several quadrants, yet a much larger lesion may occupy two entire quadrants. Both these lesions would have been considered involving two quadrants, but they may represent two very different levels of neoplasia. Assessment of lesion area would be a more accurate method to determine lesion size. Investigators at the National Cancer Institute are planning to evaluate ALTS lesions based on measurement of surface area. It also is extremely difficult to determine the exact extent of lesions when adjoining or embedded within areas of acetowhite immature metaplasia. Consequently, young women with a large cervical transformation zone and cervical lesion are exceedingly more difficult to evaluate. The ALTS population was primarily represented by younger women who may have made discrimination more challenging for reviewers compared with appraisal of older women.

Our assessment of interobserver agreement of digitized cervical images is the first involving a colposcopy quality control group participating in a large, prospective multisite cervical cytology triage study. The reported type of colposcopy quality control was complex and sophisticated at the time. However, the results were disappointingly

poor. Pretrial standardization among quality control reviewers could have a positive impact on consistency outcomes [14]. Pending detailed ALTS image analyses by the National Cancer Institute and the American Society for Colposcopy and Cervical Pathology may help standardize our assessment skills and improve levels of agreement in the future. The availability of newer digital image technology, including streaming video segments, also may enhance concordance.

Acknowledgments

Supported by a Department of Health and Human Services contract (CN-55158) with the National Cancer Institute, National Institutes of Health, Bethesda, MD. DenVu (Tucson, AZ) provided digital colposcopy systems and customized software.

AFFILIATIONS OF THE ALTS GROUP

National Cancer Institute, Bethesda, MD

D. Solomon, Project Officer
M. Schiffman, Co-Project Officer
R. Tarone, Statistician

Clinical Centers

University of Alabama at Birmingham, Birmingham, AL

E. E. Partridge, Principal Investigator
L. Kilgore, Co-Principal Investigator
S. Hester, Study Manager

University of Oklahoma, Oklahoma City, OK

J. L. Walker, Principal Investigator
G. A. Johnson, Co-Principal Investigator
A. Yadack, Study Manager

Magee-Womens Hospital of the University of Pittsburgh Medical Center Health System, Pittsburgh, PA

R. S. Guido, Principal Investigator
K. McIntyre-Seltman, Co-Principal Investigator
R. P. Edwards, Investigator
J. Gruss, Study Manager

University of Washington, Seattle, WA

N. B. Kiviat, Co-Principal Investigator
L. Koutsky, Co-Principal Investigator
C. Mao, Investigator
J. M. Haug, Study Manager

Colposcopy Quality Control Group

D. Ferris, Principal Investigator, Medical College of Georgia, Augusta, GA
J. T. Cox, Co-Investigator, University of California at Santa Barbara, Santa Barbara, CA
L. Burke, Co-Investigator, Beth Israel Deaconess Medical Center Hospital, Boston, MA

HPV Quality Control Group

C. M. Wheeler, Principal Investigator, University of New Mexico Health Sciences Center, Albuquerque, NM
C. Peyton-Goodall, Lab Manager, University of New Mexico Health Sciences Center, Albuquerque, NM
M. M. Manos, Co-Investigator, Kaiser Permanente, Oakland, CA

Pathology Quality Control Group

R. J. Kurman, Principal Investigator, Johns Hopkins Hospital, Baltimore, MD
D. L. Rosenthal, Co-Investigator, Johns Hopkins Hospital, Baltimore, MD
M. E. Sherman, Co-Investigator, Johns Hopkins Hospital, Baltimore, MD
M. H. Stoler, Co-Investigator, University of Virginia Health Science Center, Charlottesville, VA

Cost Utility Analysis Group

D. M. Harper, Investigator, Dartmouth Hitchcock Medical Center, Lebanon, NH

Westat, Coordinating Unit, Rockville, MD

J. Rosenthal, Project Director
M. Dunn, Data Management Team Leader
J. Quarantillo, Senior Systems Analyst
D. Robinson, Clinical Center Coordinator

Digene Corporation, Gaithersburg, MD

A. T. Lorincz, Senior Scientific Officer

Information Management Services, Inc, Silver Spring, MD

B. Kramer, Senior Clinical Data Manager

REFERENCES

1. Spitzer M, Apgar BS, Brotzman GL, Krumholz BA. Residency training in colposcopy: a survey of program directors in obstetrics and gynecology and family practice. *Am J Obstet Gynecol* 2001;185:507–13.

2. Ferris DG, Miller JA. Colposcopy practice and training in family practice residency programs. *J Am Board Fam Pract* 1992;5:153–6.

3. Ferris DG, Miller JA, Wagner P, Walaitis E, Lawler FH. Clinical decision-making following abnormal Papanicolaou smear reports. *Family Pract Res J* 1993;13:343–53.

4. Ferris DG, Miller JA. Colposcopic accuracy in a residency training program: defining competency and proficiency. *J Fam Pract* 1993;36:515–20.

5. Benedet JL, Anderson GH, Matisic JP, Miller JA. A quality control program for colposcopic practice. *Obstet Gynecol* 1991;78:872–5.

6. Ferris DG, Spitzer M, Werner C, Dickman ED, Shiver RL. Colposcopy quality control for clinical trials: the positive effects from brief, intensive educational intervention. *J Lower Genital Tract Dis* 2002;6:11–5.

7. Schiffman M, Adrianza ME. ASCUS-LSIL triage study (ALTS): design, method, and characteristics of trial participants. *Acta Cytol* 2000;44:726–42.

8. Ferris DG, Cox JT, Burke L, Litaker MS, Harper DM, Campion MJ, et al. Colposcopy quality control: establishing colposcopy criterion standards for the National Cancer Institute ALTS trial using cervigrams. *J Lower Gen Tract Dis* 1998;2: 195–203.

9. Ferris DG, Litaker MS, the ALTS group. Colposcopy quality control by remote review of digitized cervical images. *Am J Obstet Gynecol.* In press.

10. Weatherholt JW. Colposcopy recognition award examination announcement. *Colposcopist* 1994;24:3.

11. Reid R, Scalzi P. Genital warts and cervical cancer. VII. An improved colposcopic index for differentiating benign papillomaviral infections from high-grade cervical intraepithelial neoplasia. *Am J Obstet Gynecol* 1985;153:611–8.

12. Ferris DG, Greenberg, MD. Reid's colposcopic index. *J Fam Pract* 1994;39:65–70.

13. Stoler MH, Schiffman MH. Interobserver reproducibility of cervical cytologic and histologic interpretations. *JAMA* 2001;285:1500–5.

14. DeVet HCW, Koudstaal J, Kwee WS, Willebrand D, Arends JW. Efforts to improve interobsever agreement in histopathological grading. *J Clin Epidemiol* 1995;48:869–73.

15. McCluggage WG, Walsh MY, Thornton CM, Hamilton PW, Date A, Caughley LM, et al. Inter- and intraobserver variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system. *Br J Obstet Gynaecol* 1998;105:206–10.

16. Sellors JW, Nieminen P, Vesterinen E, Paavonen J. Observer variability in the scoring of colpophotographs. *Obstet Gynecol* 1990;76:1006–8.

17. Hopman EH, Voorhorst FJ, Kenemans P, Meyer CJL, Helmerhorst TJM. Observer agreement on interpreting colposcopic images of CIN. *Gynecol Oncol* 1995;58:206–9.

18. Etherington IJ, Luesley DM, Shafi MI, Dunn J, Hiller L, Jordan JA. Observer variability among colposcopists from the West Midlands region. *Br J Obstet Gynaecol* 1997;104: 1380–4.

19. Ferris DG, Macfee MS, Miller JA, Litaker MS, Crawley D, Watson D. The efficacy of telecolposcopy compared with traditional colposcopy. *Obstet Gynecol* 2002;99:248–54.

20. Ferris DG, Bishai DM, Macfee MS, Litaker MS, Dickman ED, Miller JA. Telemedicine network telecolposcopy compared with computer-based telecolposcopy. *J Lower Genital Tract Dis* 2004;8:94–101.